Semantic, Lexical, and Geographic Cues are used in Geographic Fluency

Janelle Szary (jszary@indiana.edu)
Michael N. Jones (jonesmn@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. 10th Street
Bloomington, IN 47405 USA

Abstract

Semantic fluency tasks have increasingly been used to probe the structure of human memory, adopting methodologies from the ecological foraging literature to describe memory as a trajectory through semantic space. Clusters of semantically related items are often produced together, and the transitions between these clusters of semantically related items are consistent with theories of optimal foraging, where the search process exhibits a balance between exploration and exploitation behaviors (Hills, Jones, & Todd, 2012). Here, we use a semantic fluency memory task in which subjects recall geographic locations. For each pairwise transition, we measure temporal, geographic, semantic, lexical, and phonetic distances. In general, the dimensions are loosely but reliably correlated with each other. Segmentation of the retrieval sequence into patches supports the notion that subjects strategically leave patches as within-patch resources diminish, but also suggests that subjects may shift their attention between different sources of information, perhaps reflecting dynamically changing patch definitions.

Keywords: Memory search; Semantic fluency; Optimal foraging; Spatial search

Introduction

In general, effective search strategies require a balance of both exploration and exploitation. The searcher, whether it be a bumblebee searching for nectar, or a child searching for hidden Easter eggs, needs to know when to stop looking in one place and start looking in another. Too much exploration, in the later example, would be a situation in which the child runs wildly around the yard, but never stops to looks behind any leaves (where the Easter eggs would be hidden); Too much exploitation, on the other hand, might be a situation in which the child, finding one egg hidden in the rose bushes, spends the afternoon meticulously sorting through the roses for more eggs, and fails to search the rest of the yard. Although this example is exaggerated, it illustrates the importance of finding an appropriate balance between the two strategies: At either extreme, the child will do very poorly in an Easter egg hunt. A successful search process relies on appropriately modulating between exploration and exploitation.

The same principle applies to search in the cognitive domain: At one end of the extreme, a lack of focus would make it hard to retrieve relevant information, and, at the other end of the extreme, perseveration on one piece of information would also be ineffective. The parallel between search in physical space and cognitive space is not coincidental. Significant evidence suggests that the neural and molecular processes that evolved to govern search in physical space have been exapted to control goal-directed behaviors in other, cognitive modalities (reviewed in Hills, 2006). For example, the dopaminergic pathways of the basal ganglia control both cognitive attention and movement, and increases in dopamine have been associated with exploitation of resources and highly focused behaviors, while reductions in dopamine have been associated with exploratory or inattentive behaviors (see Hills, 2006). The exaptation hypothesis has led to the idea that there are general cognitive search mechanisms that control goal-directed search behaviors in both internal and external spaces (Hills, Todd, & Goldstone, 2008). As evidence for this claim, Hills et al. (2008) demonstrated that priming certain physical search strategies can influence cognitive search behavior.

Further support for the exaptation hypothesis comes from the successful application of models from the animal foraging literature to cognitive search behavior. For example, in the foraging literature, a search process is considered optimal when it follows the marginal value theorem (MVT; Charnov, 1976), in which a forager should continue exploiting a patch of resources while it continues to provide relative rewards, but should leave and switch to a new patch once the rate of rewards (for the given patch) drops below a long-term average. Hills et al. (2012) used the semantic fluency task, in which subjects are asked to name as many items from a given category as possible, to show that cognitive search is well described by MVT, where patches are clusters of semantically related items. As a subject's relative success in a given semantic cluster decreases, the subject is more likely to switch to a new semantic cluster. Models of optimal foraging in semantic memory based on MVT have been successful in matching human performance, but they rely on the construction of appropriate semantic spaces and patches, which can be done in different ways with differing results (see discussion in Hills et al., 2015; Abbott, Austerweil, & Griffiths, 2015).

Anecdotally, the similarity between physical and cognitive search is supported by the fact that we often describe our internal, cognitive representations of information as networks or maps (Steyvers & Tenenbaum, 2005; Tolman, 1948). Further, there is growing support for the embodied cognition perspective that cognitive knowledge is grounded in physical space. Montez, Thompson, and Kello (2015) asked participants to spatially organize a set of items produced during a previous semantic fluency task, and found that the spatial distances correlated with the previously observed temporal distances: Subjects spatially organized items the same way they had previously organized them during semantic retrieval. Louwerse and colleagues (Recchia & Louwerse, 2014; Louwerse & Zwaan, 2009) have shown that from the statistics of language we can extract a great deal of information, such as the locations and sizes of cities.

Using an extended semantic fluency task where subjects spent twenty minutes retrieving the names of cities and towns from their home state of California, Szary, Kello, and Dale (2015) showed that the temporal structure of retrieval sequences captured the geographic structure of cities in physical space. Although the task is known as a 'semantic fluency' task, previously we could only infer semantic information from the timing of the retrieval process. Here, we build upon the study presented in Szary et al. (2015) by computing semantic, lexical, and phonetic measures to compare to the existing geographic and temporal measures. The new analyses allow us to formulate new research questions. Consider the hypothesis that internal search is exapted from external search, and relies on a balance between exploration and exploitation. Because physical space has a finite number of dimensions, it is quite easy to measure exploration versus exploitation behaviors in external search. By comparison, it is less straightforward how one should characterize exploration or exploitation behaviors in the high-dimensional cognitive space of an internal search. Here, we re-investigate semantic fluency data along a number of dimensions: temporal, geographic, semantic, lexical, and phonetic. In this exploratory study, we show that retrieval sequences do indeed contain structure along each of these dimensions. That is, when remembering geographic information, people do, indeed, use the embodied physical structure of the information, but we can also measure their use of semantic, lexical, and even phonetic information. With regards to the optimal foraging perspective, these metrics may reflect different dimensions in which patches exist, highlighting a number of open questions for future work, as discussed in the conclusion.

Methods

Experimental Procedure

Participants Participants were recruited from a subject pool of University of California, Merced undergraduate students who participated for course credit (4 male, 8 female; mean age = 19.92 years, *SD* = 1.08 years), and reported being native or proficient English speakers. All but two participants reported living in California for their whole lives, while the other two reported living in California for the majority of their lives (15 of 19 years, and 16 of 18 years, respectively). Subjects were comfortably seated by themselves at a table in small experiment room, and wore Shure microphone headsets. Speech was collected using an M-Audio MobilePre recording interface and Audacity software.

Task Subjects completed two recall tasks, presented in counterbalanced order, each of which lasted for twenty minutes. In one task subjects recalled items from the category of cities and towns in California, and in the other they recalled from the category of all animals. For the purposes of the present paper, we discuss only results from the category of cities and towns in California. Subjects were given the following instructions: "Your goal [is] to think of as many items from [the] category as you can. When you think of an item, just say it out loud. You can be as specific or as general as you wish. For example, if the category were *Food* you could say 'Fruit', and you could also say 'Orange' or 'Mandarin Orange'. But keep in mind that your goal is to recall as many different items as possible. If you are unsure if an item does or does not belong to the category, just say it anyhow, don't spend time worrying about whether something counts or not," (adapted from Rhodes & Turvey, 2007).

Audio Transcription The speech recordings were loaded into Praat audio analysis software for annotation, and were transcribed as in Szary et al. (2015). That is, each recalled item was transcribed, and onset times of items were marked. Repeated items were removed. Incorrect items ("Reno", which is in Nevada, not California), geographic landmarks ("Monterey Bay" bay, "Sierra Nevadas" mountains), and non-specific areas ("Bay Area", which refers to several locations around the San Francisco Bay) were removed. Pronunciation errors ("Rancho Cucamongo" instead of the correct "Rancho Cucamonga") and common abbreviations ("L.A." instead of the official "Los Angeles") were corrected. Districts, neighborhoods, planned communities and census-designated areas with names recognized by the U.S. Geological Survey (e.g. "Hollywood", "Downieville"; *United States Board on Geographic Names*, 2016) were retained.

Measures

For each dataset, inter-retrieval intervals (*IRIs*) are measured as the amount of time (in milliseconds) between consecutive recall events $city_i$ and $city_{i+1}$, and represent the temporal distances we observed. For each transition ($city_i$ to $city_{i+1}$), we compute additional distances using geographic, semantic, phonetic, and lexical measures.

Geographic Distance The latitudinal and longitudinal coordinates for most cities were provided by the world.cities dataset, in the maps package for the R programming language (Becker, Wilks, Brownrigg, Minka, & Deckmyn, 2016). Missing values were added by hand, and were retrieved from Wikimedia's GeoHack tool (*Wikimedia Tool Labs GeoHack*, 2016). Geographic distances (*GDs*) are measured as the number of miles between consecutively recalled cities. GDs are calculated using the Haversine formula, which gives the great-circle distance between two points on an sphere (Sinnott, 1984).

Semantic Distance Semantic similarity was calculated using vector representations obtained by a Word2Vec model from the gensim package for Python (Řehůřek & Sojka, 2010), which was trained on the full English Wikipedia dataset, available from *dumps.wikimedia.org/enwiki*, 2016. The cosine between the vectors representing two city names is taken as a measure of the semantic distance between those cities.

Lexical Distance The standard Levenshtein distance, which measures the number of edits required to transform one

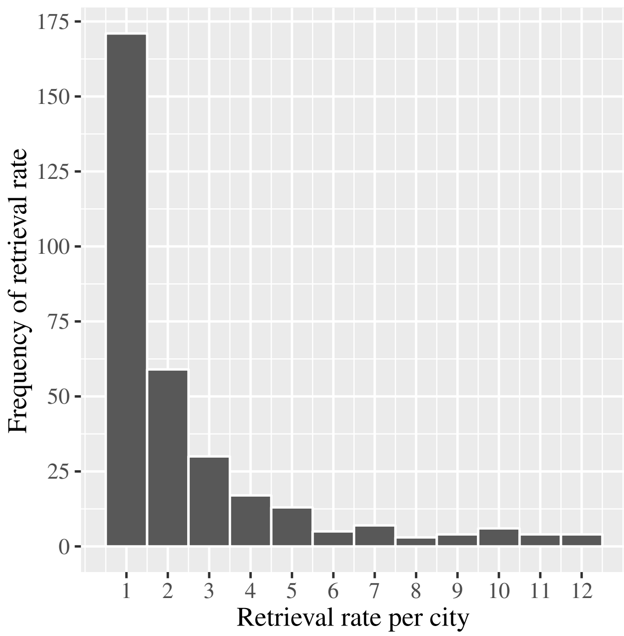

Figure 1: The number of cities that has each retrieval rate.

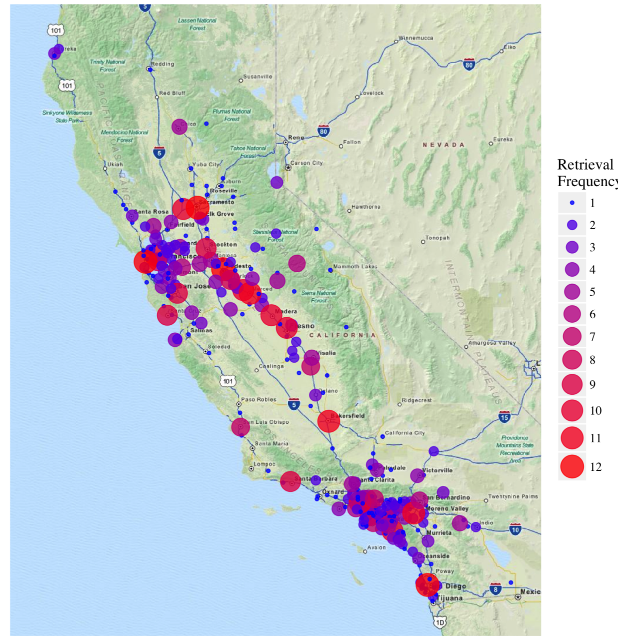

Figure 2: City locations and their retrieval rates.

string into another, is used as a measure of lexical distance.

Phonological Distance Finally, as an exploratory measure of phonological difference, we converted city names into their four-digit soundex codes, which are approximate representations of how words sound. We then compute the Levenshtein distance between these codes, which serves as our measure of phonological distance.

Results

Across all participants, a total of 323 unique city names were retrieved. On average, each participant retrieved 67 cities ($SD = 28$). Many of the cities ($n = 171$) were retrieved only once, and only 4 cities were retrieved by all 12 subjects. Figure 1 shows the number of cities with each retrieval rate. To visualize the structure in these highly variable retrieval rates, Figure 2 shows the locations of retrieved cities along with their retrieval frequencies.

Distance Measures

Across all transitions in our dataset, our distance measures tended to be subtly correlated. Replicating the findings from Szary et al., 2015, geographic similarity was positively correlated with temporal proximity across all observed pairwise transitions, as measured by $-IRI$s, $r(787) = 0.16$, $p < 0.001$. In addition, we found that geographic similarity captured semantic similarity, $r(787) = 0.13$, $p < 0.001$. Figure 3 shows the positive correlations between geographic distance measures and both semantic and temporal distance measures. As evident in this graph, semantic and temporal distances are also closely related, with $r(787) = 0.17$, $p < 0.001$. While temporal distance has slight positive correlations to both lexical and phonological distance, $r(787) = 0.12$, $p < 0.001$ and $r(787) = 0.08$, $p = 0.02$, respectively, neither lexical nor phonological distances are correlated with geographic distance. Finally, although lexical and phonological distances are correlated, $r(787) = 0.21$, $p < 0.001$, they have different relationships to the semantic distance measure. Semantic and phonological distances are positively correlated, $r(787) = 0.18$, $p < 0.001$, but semantic and lexical distances actually show a slight negative correlation, $r(787) = -0.09$, $p < 0.01$.

Observed Versus Randomized Distances To make sure that the distance measures differ from those that would be observed by chance, we shuffled each subject's retrieval sequence 100 times, and compared this to simulated data. For each metric, distances were greater for the shuffled data as compared to the original data, as seen in Figure 4. Geographic distances were greater in the shuffled data ($M = 172.7$ miles) than the observed data ($M = 115.0$ miles), $t(813) = -12.76$, $p < 0.001$. Semantic distances were also greater in the shuffled data ($M = 0.71$) than the observed data ($M = 0.63$), $t(798) = -12.9$, $p < -.001$. Lexical distances were greater in the shuffled data ($M = 8.15$) as compared to the observed data ($M = 8.39$), $t(836) = -2.81$, $p < 0.01$. Finally, phonetic distances were greater in the shuffled data ($M = 3.26$) as compared to the observed data ($M = 3.35$), $t(804) = -2.94$, $p < 0.01$.

Patch Transitions

Kernel density estimation was used to determine a segmentation threshold (at the first local minimum) for each participant's IRI sequence ($M = 28$, $SD = 10$ seconds). This thresh-

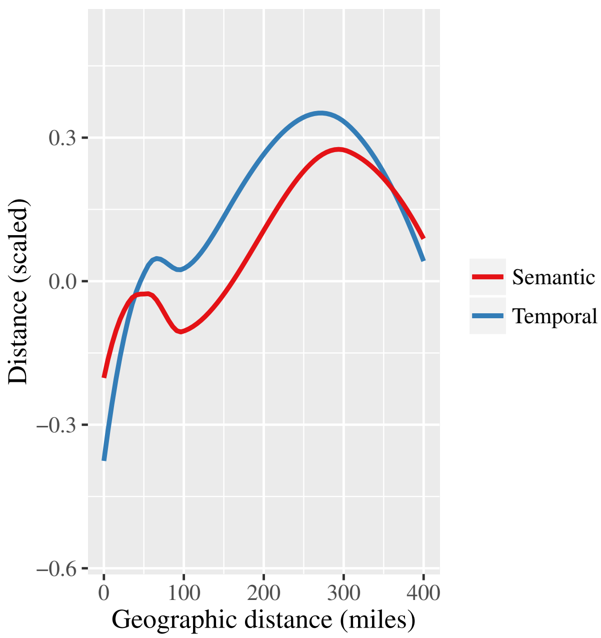

Figure 3: Semantic and temporal distances as a function of geographic distance. For visibility (but not in the analyses), semantic and temporal ranges are scaled and normalized, and geographic extremes are removed. A loess smoothing curve is fit to remaining datapoints to show trends.

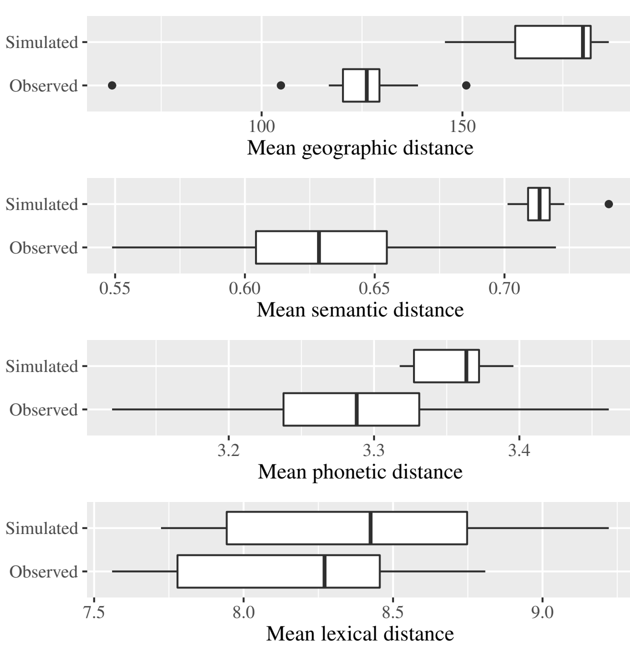

Figure 4: Boxplots showing the differences in mean distances between the simulated and observed datasets.

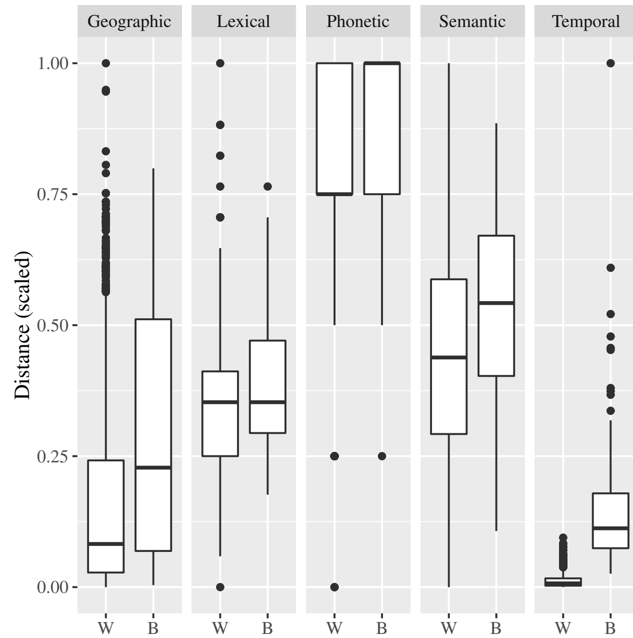

Figure 5: Boxplot of group means comparing pairwise transitions that occurred within patches ("W") to pairwise transitions that occurred between patches ("B"). Distance measures are scaled to share a common axis for illustration purposes only.

old was used to segment retrieved items into patches. Our datasets had an average of 11.4 ($SD = 4.7$) patches. While most patches were small (median size = 2), some were quite large (maximum size = 51; $M = 5.8$, $SD = 9.2$).

Figure 5 compares pairwise transitions that occurred within patches to the pairwise transitions that occurred at patch boundaries (switches). Mean semantic and geographic distances were both significantly higher between patches (at the transition points) as compared to within patches. For semantic distances, $M_{between} = 0.71$ and $M_{within} = 0.62$, $t(211) = 5.80$, $p < 0.001$; For geographic distances $M_{between} = 162.2$ miles and $M_{within} = 105.5$ miles, $t(181) = 4.56$, $p < 0.001$. Lexical distance was slightly but significantly higher between patches ($M = 8.68$) as compared to within patches ($M = 8.05$), $t(194) = 2.94$, $p < 0.005$. Finally, phonetic distance was also slightly higher between patches ($M = 3.44$) as compared to within patches ($M = 3.23$), $t(235) = 3.10$, $p < 0.005$.

Figure 6 shows this same data, but instead of averaging across all within-patch distances, considers the transitions immediately preceding patch switches. Specifically, it shows averages for the switches (index = 0) and the 9 transitions preceding switches. While patch switches have higher average distances on each metric, Figure 6 suggests that there may be patterns leading up to patch switches, perhaps reflecting the depletion of resources available through a given metric.

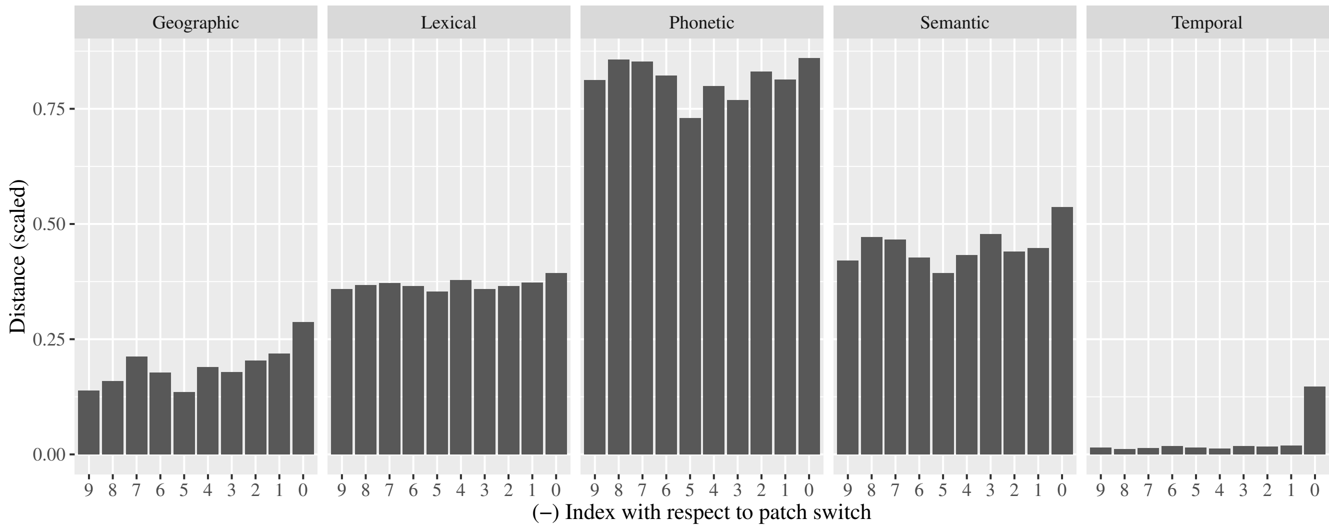

Figure 6: Bar heights show scaled averages for each distance metric at the 9 transitions preceding a patch switch, and the switch itself (index = 0).

Shifting Strategies

Although we focus primarily on summary statistics in this paper, our larger goal is to explore how different cues (geographic, semantic, lexical, phonetic) interact dynamically during retrieval processes. Although a detailed quantitative analysis is beyond the scope of the present paper, Figure shows the evolution of all 4 distance metrics over the course of the trial for each subject, in service of our larger goal.

While there are some similarities across participants (such as geographic distance being relatively lower than other measures), there are also striking differences in the way the measures unfold over time for the different participants. This observation is consistent with the idea that distances may reflect the cueing strategies underlying the search process, where geographic cues would lead a subject to retrieve geographically proximal items, resulting in low geographic distances. Consider, as a simple example, the phonetic and semantic distances for Subject 12. Over the course of the task, semantic distances show a relative increase, while phonetic distances show a relative decrease. From the perspective that our distance metrics (in some loose way) reflect the cueing strategies at work, we see that Subject 12 initially uses a semantic cueing strategy, thinking of items that are semantically related to one another (perhaps listing affluent cities with colleges: "La Jolla... Berkeley... Stanford"). But, as semantically related resources are depleted, the subject switches to a phonetic strategy (perhaps "Stanford... Stanton... Winton").

General Discussion

Overall, most of our distance metrics were positively correlated with each other. Temporal distances were correlated with geographic, semantic, lexical, and phonetic distances. Additionally, geographic distances were correlated with semantic distances, but not with lexical or phonetic distances. Semantic distances were correlated positively with phonetic distances, but negatively with lexical distances. Still, lexical and phonetic distances were positively correlated.

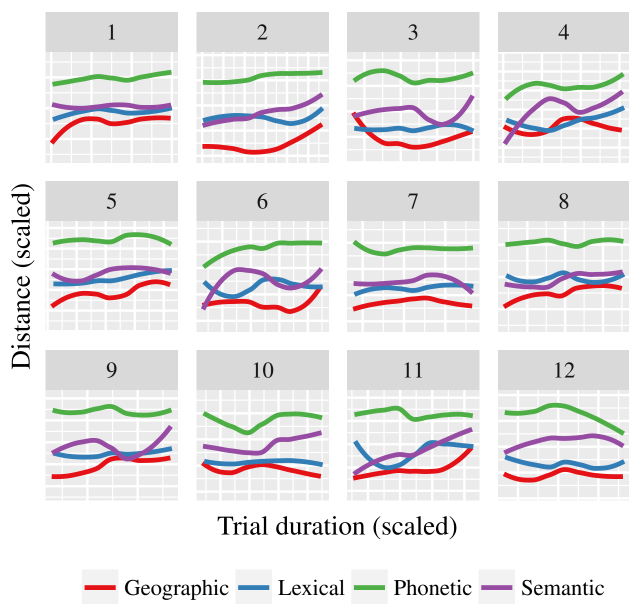

Figure 7: Loess curve fits to each metric (scaled) for each subject.

To test whether there is geographic, semantic, phonetic, or lexical structure in the sequence of retrieved items beyond what would be observed by chance, we compared distances from the observed sequences to distances from sequences of the same items in shuffled orders, as in Szary et al. (2015).

For each of our distance measures, simulated datasets with randomized orders showed greater distances. That is, there is structure in the sequence of retrieved items beyond what would be expected for a random sampling of items, or if a given distance metric were unrelated. This suggests that subjects use each source of information to cue their retrieval processes. In the actual datasets, but not shuffled datasets, pairwise transitions include items that are closely related (on one or more metric).

We also find that each distance metric spikes at the transitions between patches, and tends to increase in the transitions leading up to patch transitions (seeFigure 6). While a quantitative analysis is beyond the scope of the current paper, this finding is consistent with the notion that semantic search utilizes an optimal foraging policy. A hypothesis consistent with optimal foraging would predict that, as a forager retrieves the resources within a given semantic patch, the availability of additional resources becomes depleted, which may be reflected by increases in the the amount of time to find additional items in the transitions preceding patch switches. Here, we use an automatic clustering algorithm applied to the temporal sequence to define patches. While temporal distances are used to define the patch switches, it is difficult to draw conclusions about how they may effect those switches. Visual inspection of the other metrics, however, is consistent with the idea that patches are becoming increasingly sparse (across multiple dimensions) in the time leading up to the decision to switch.

For the argument that search is a generalized cognitive process which relies on the appropriate balance between exploration and exploitation, the word *appropriate* is key. What makes the balance appropriate is known to change depending on search context, such as whether you're searching for randomly distributed or highly clustered resources, but it can also change dynamically over the course of a search process, such as when the resources in a given patch have been depleted. An open question, then, is how the different dimensions of relatedness (geographic, semantic...) interact in what we think of as 'patches'. In order to accurately characterize the line between exploration and exploitation in cognitive space, future work must address how these information sources combine in patches. For example, how does a forager incorporate multidimensional information? Are the patches themselves multidimensional? Or do we move, orthogonally, through cue-strategies and information patches, respectively? As increasing evidence suggests that search strategies are a domain-general process, across a number of cognitive contexts (Hills, Todd, & Goldstone, 2010; Rhodes & Turvey, 2007), a complete understanding of how we leverage different cues to sift through the information in semantic memory becomes increasingly important.

Acknowledgments

Thank you to Fatemeh Torabi Asr and Rick Dale for assistance with data collection and analysis.

References

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psychological Review*, *122*(3), 558-569.

Becker, R. A., Wilks, A. R., Brownrigg, R., Minka, T. P., & Deckmyn, A. (2016). *maps: Draw geographical maps* (R package version 3.1.0 ed.). Retrieved 2016, from `https://CRAN.R-project.org/package=maps`

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*(2), 129-136.

dumps.wikimedia.org/enwiki. (2016). Retrieved from `https://dumps.wikimedia.org/enwiki/latest/`

Hills, T. (2006). Animal foraging and the evolution of goal-directed cognition. *Cognitive Science*, *30*, 3-41.

Hills, T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431-440.

Hills, T., Todd, P. M., & Goldstone, R. L. (2008). Search in external and internal spaces: Evidence for generalized cognitive search processes. *Psychological Science*, *19*(8), 802-808.

Hills, T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., & Group, T. C. S. R. (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences*, *19*(1), 46-54.

Louwerse, M. M., & Zwaan, R. A. (2009). Language encodes geographical information. *Cognitive Science*, *33*, 51-73.

Montez, P., Thompson, G., & Kello, C. T. (2015). The role of semantic clustering in optimal memory foraging. *Cognitive Science*, *39*(8), 1925-1939.

Recchia, G., & Louwerse, M. (2014). Grounding the ungrounded: Estimating locations of unknown place names from linguistic associations and grounded representations. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (p. 1270-1275). Austin, TX: Cognitive Science Society.

Řehůřek, R., & Sojka, P. (2010, May 22). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50). Valletta, Malta: ELRA. (`http://is.muni.cz/publication/884893/en`)

Rhodes, T., & Turvey, M. T. (2007). Human memory retrieval as lévy foraging. *Physica A: Satistical Mechanics and its Applications*, *385*(1), 255-260.

Sinnott, R. W. (1984). Virtues of the haversine. *Sky and Telescope*, *68*(2), 159.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*, 41-78.

Szary, J., Kello, C., & Dale, R. (2015). Memory foraging in a spatial domain. In *Proceedings of the 37th annual conference of the cognitive science society.* Austin, TX: Cognitive Science Society.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189-208.

United states board on geographic names. (2016). Retrieved from `geonames.usgs.gov`

Wikimedia tool labs geohack. (2016). Retrieved from `https://www.mediawiki.org/wiki/GeoHack`